



BERT代码从零解读

NLP入门到放弃-DASOU

BERT代码最核心的一点是：MLM损失函数的计算

15% 词汇

- **8份真正被mask**
- **1份被随机替换**
- **1份保持不变**

计算MLM损失的时候使用哪部分？

后台回复【代码解析】



扫码关注微信公众号

文章周更

知识分享

一起进步

求关注，求点赞，求一切！！

两个重点： 第一只使用了TRM的编码部分

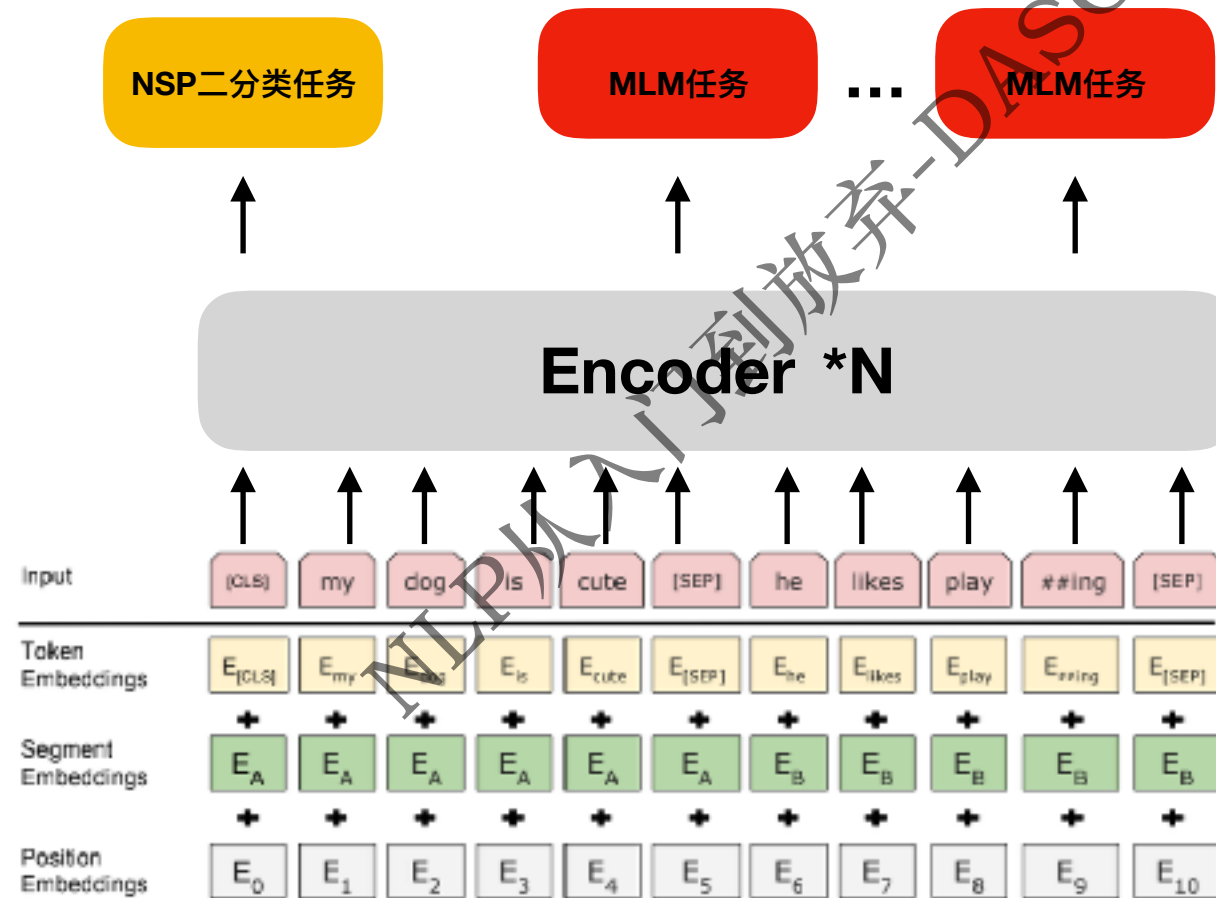
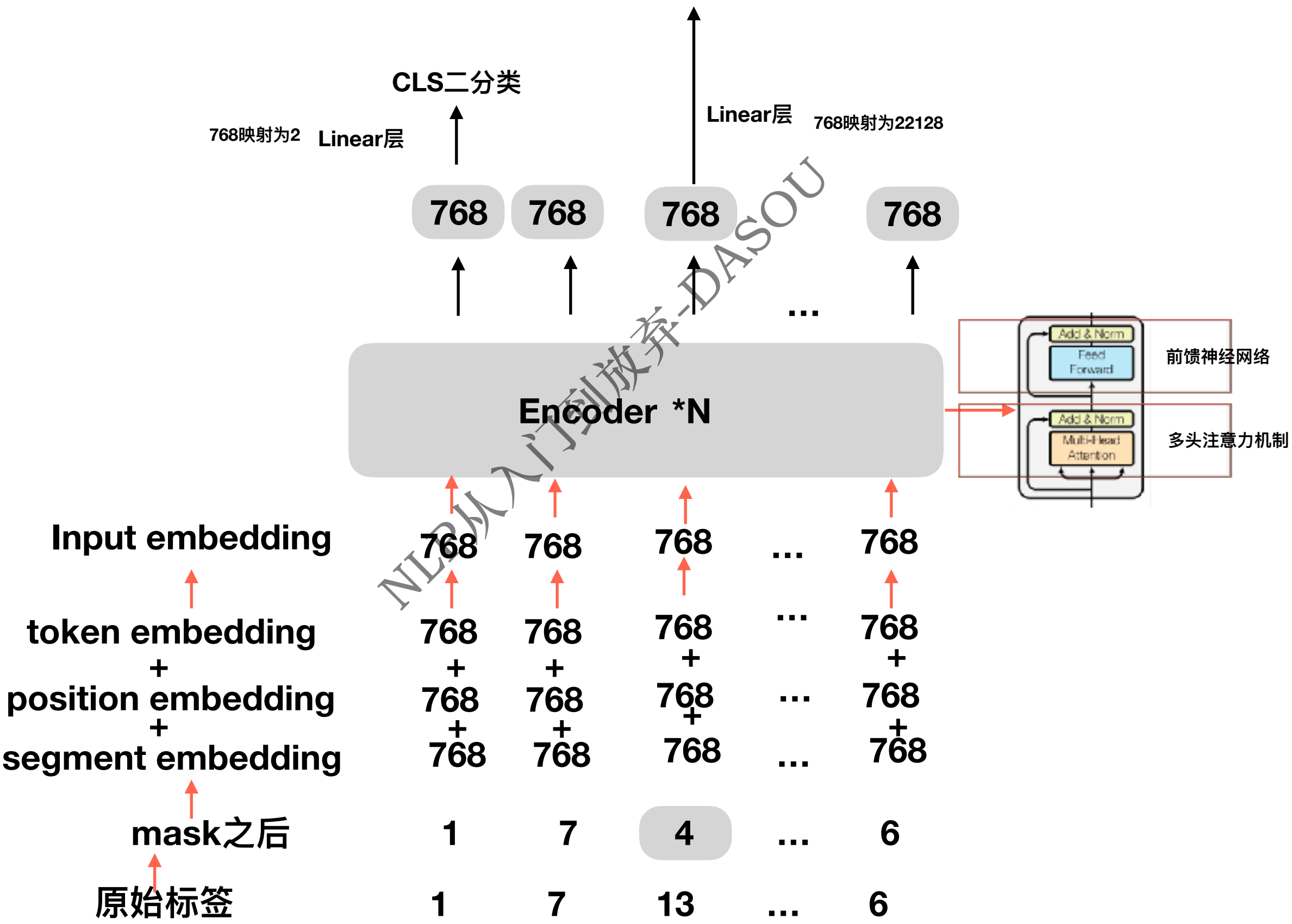
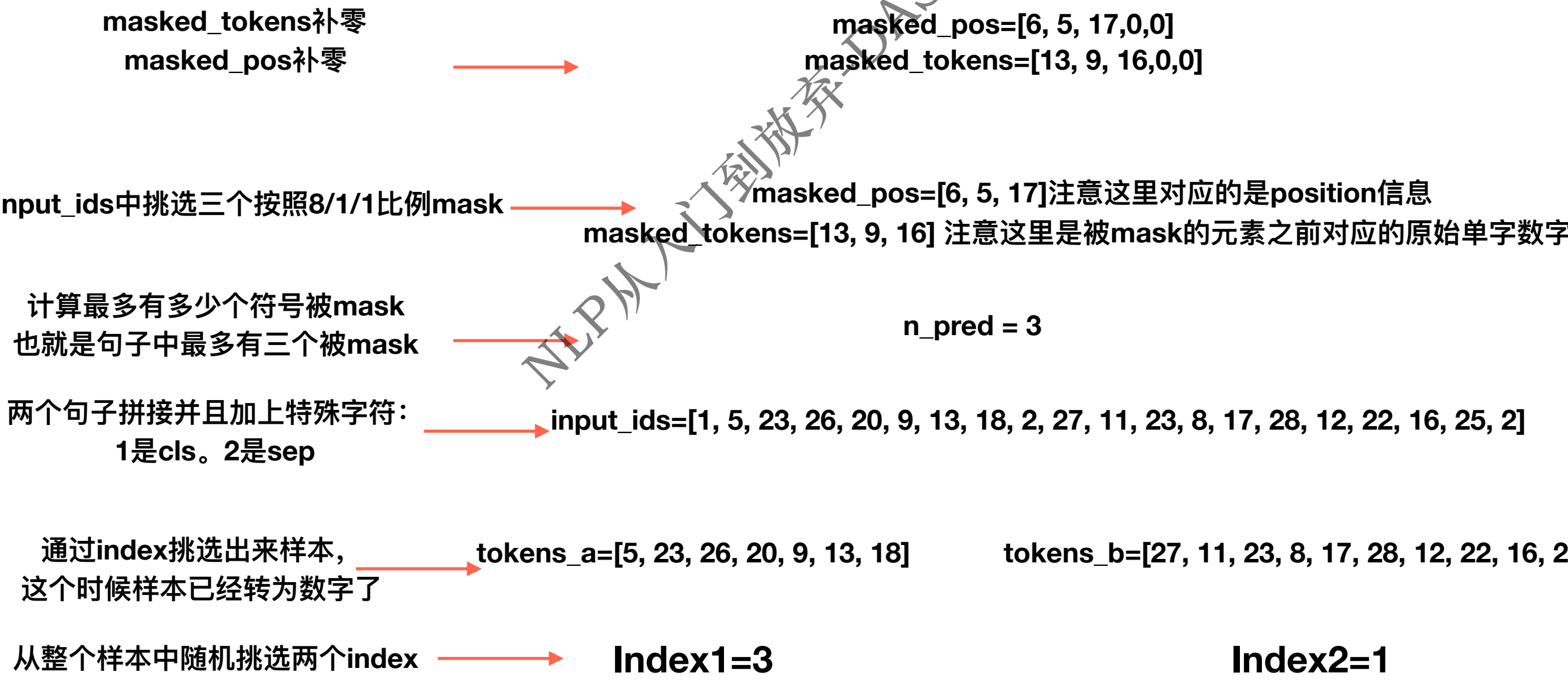


Figure 2: BERT input representation. The input embeddings are the sum of the token embeddings, the segmentation embeddings and the position embeddings.

22128词表大小-softmax



最大问题来了：为啥是？补零，补别的可以吗



最大问题来了：为啥是？补零，补别的可以吗

```
class CrossEntropyLoss(WeightedLoss):  
  
    def __init__(self, weight: Optional[Tensor] = None, size_average=None, ignore_index: int = -100,  
                 reduce=None, reduction: str = 'mean') -> None:  
        super(CrossEntropyLoss, self).__init__(weight, size_average, reduce, reduction)  
        self.ignore_index = ignore_index  
  
    def forward(self, input: Tensor, target: Tensor) -> Tensor:  
        return F.cross_entropy(input, target, weight=self.weight,  
                                ignore_index=self.ignore_index, reduction=self.reduction)
```

有一个ignore_index的参数：忽略掉某一个类别；补零的话，忽略掉GT为0的